Dai, Q., Bray, M., Zhuo, L., Islam, T., & Han, D. (2017). A Scheme for Rain Gauge Network Design Based on Remotely Sensed Rainfall Measurements. *Journal of Hydrometeorology*, *18*(2), 363. https://doi.org/10.1175/JHM-D-16-0136.1

A scheme for raingauge network design based on remotely-sensed rainfall measurements

Qiang Dai[1,2,3,*], Michaela Bray[4], Lu Zhuo[2], Tanvir Islam[5], Dawei Han[2]

[1]Key Laboratory of VGE of Ministry of Education, Nanjing Normal University, Nanjing, China
[2]WEMRC, Department of Civil Engineering, University of Bristol, Bristol, UK
[3]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China
[4]Hydro-Environmental Research Center, Cardiff University, Cardiff, UK
[5]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[*]Correspondence: q.dai@njnu.edu.cn or dqgis@hotmail.com

Abstract:

We have been facing a remarkable decline in the number of raingauges in many areas of the world, as a compromise to the expensive cost of operating and maintaining raingauges. The question of how to effectively deploy new or remove current raingauges in order to create optimal rainfall information is becoming more and more important. On the other hand, larger-scaled remotely-sensed rainfall measurements, although poorer quality compared with traditional raingauge rainfall measurements, provide an insight into the local storm characteristics, which traditional methods for designing a raingauge network sort to seek. Based on these facts, this study proposes a new methodology for raingauge network design using remotely-sensed rainfall data set, which aims to explore how many gauges are essential and where they should be placed. Principal component analysis (PCA) is used to analyse the redundancy of the radar grids network and determine the number of raingauges while the potential locations are determined by cluster analysis (CA) selection. The proposed methodology has been performed on 373 different storm events measured by a weather radar grids network, and compared against an existing dense raingauge network in Southwest England. Due to the simple structure, the proposed scheme could be easily implemented

in other study areas. This study provides a new insight into raingauge network design, which is also a preliminary attempt of using remotely-sensed data to solve the traditional raingauge problems.

Keywords: radar rainfall; raingauge network design; principal component analysis; remotely-sensed rainfall

1. Introduction

Rain is a major component of the water cycle on the earth, which is a significant issue in many scientific fields such as ecosystems, agriculture and water environment. Methods for measuring rainfall need to take into account its mutability in order to minimise uncertainty and the errors found within the recorded data. The use of raingauges is one of the oldest and most common methods employed in the world for measuring rainfall. Nowadays, raingauge rainfall is a vital source of information used for the calibration of remotely-sensed rainfall and verification of numerical weather model rainfall products. However, hydrologists and meteorologists are facing a dilemma of raingauge management. Additionally raingauges are extremely desired in many regions to improve the local rainfall quality, especially in ungauged catchments, while on the other hand, expensive cost of operating and maintaining of raingauges leads to the remarkable shut down of available raingauges. A decline of roughly 50% of raingauges numbers has occurred in the period 1989–2006 in Europe, South America, and Africa [*Lorenz and Kunstmann*, 2012; *Walsh*, 2012] and an approximate 50% decrease in number of valid daily reports in the period 2000-2007 for APHRODITE [*Overeem et al.*, 2013; *Yatagai et al.*, 2012]. In an effort to address this dilemma, an efficient and feasible scheme for raingauge network design is needed to capture maximum rainfall information with a minimum number of gauges.

The traditional raingauge network design is classified by two types: haphazard manner and quantitative method. The haphazard manner is generally based on numerous technical "guidelines" or "considerations". At present, there is still no standard procedure in place in most parts of the world for raingauge network design. The reason for this is due to the complexity of the problem faced by both hydrologists and meteorologists. Raingauges were distributed according to the population, in order to be close to 'observers'. This led to areas of high rainfall having relatively few gauges. Design requirements consist of determining the number of gauges, and their locations, given the frequency in time of sampling which minimises the uncertainty of rainfall estimation. Other factors also need consideration, i.e. the nature of the catchment, its topographic influences, its drainage patterns, the accessibility and suitability of proposed locations and the cost of installing and maintaining the gauges. Moreover the purpose of the network and the regional climate should be taken into consideration. The high variability of rainfall over time and space is a significant issue and is difficult to address; to typify rainfall patterns of high variability and intermittency a much denser network would be needed [*Barancourt et al.*, 1992; *Rodríguez-Iturbe and Mejía*, 1974].

Raingauge network design based on quantitative analysis attracts more attention. There are many methods employed, such as spatial correlation, Variogram Analysis and Entropy theory [*Al-Zahrani and Husain*, 1998; *Bastin et al.*, 1984; *Bogárdi et al.*, 1985; *Bradley et al.*, 2002; *Bras and Rodríguez-Iturbe*, 1985; *Krstanovic and Singh*, 1992; *Mishra and Coulibaly*, 2009; *Moore et al.*, 1999; *Pardo-Igúzquiza*, 1998; *Tsintikidis et al.*, 2011; *Volkmann et al.*, 2010; *Yang and Burn*, 1994]. Statistical techniques include Variance Reduction Algorithm, State-Space Stochastic Models and Generalized Least Squares are also adopted in numerous studies [*Bradley et al.*, 2007; *Morrissey et al.*, 1995; *Moss and Tasker*, 1991; *Shih*, 1982; *Stedinger and Tasker*, 1985; *Tasker and Moss*, 1979]. The more sophisticated techniques of raingauge network design can provide

some insight into the location of raingauges as well as the density of the network. These methods are generally based on analysing the available limited raingauge information or duplicating the learned knowledge from mature network to a pioneering area. The nature of these methods makes them difficult to implement, and often requires subjective parameter adjustments. Another shortcoming of most of the aforementioned methods is that they require exhausting all possible candidate networks to explore the optimum one.

Compared with the traditional methods that collect and excavate the limited rainfall-related information (such as rainfall spatial correlation and rainfall patterns) from inside or outside a study area, adoption of remotely-sensed rainfall measurements for raingauge network design is obviously a more direct and efficient way. Currently, remotely-sensed rainfall estimates are available in most parts of the world as some satellites have global rainfall observing ability, such as the Tropical Rainfall Measuring Mission (TRMM) and the Global Precipitation Measurement (GPM). Weather radars are also extensively installed around the world, especially in developed regions. Due to the poorer quality, these rainfall products cannot replace raingauge measurements at present, and raingauges will still be the first choice in most hydrological applications in the near future. However, these relatively inaccurate but larger-scaled rainfall measurements are an ideal data set to provide insight into the local storm characteristics, such as rainfall patterns and local topographic influences on rainfall, which are the key components that the traditional raingauge network designed methods eagerly expect to explore. Through analysing the long-term remotely-sensed rainfall data set, we can reveal these characteristics and consequently investigate the most important locations for deploying raingauges in the study area. For this reason, this study presents a new approach to raingauge network design based on a combination of principal component analysis (PCA) and variable selection criteria using weather radar rainfall data set to provide both

the optimum raingauge density and raingauge location. This scheme offers a new insight into raingauge network design and could be easily adopted and implemented in other study areas.

This paper is organized as follows. After the introduction, Section 2 illustrates the study area and data sets used. The following methodology section describes the algorithms of determining the redundancy of networks and optimum locations of gauges. Section 4 presents the outcomes of the raingauge network design and evaluations of the proposed method. The discussion section elaborates on three key issues associated with this scheme. The final conclusion section summarizes the key findings and the future work.

2. Study area and data source

The Brue Catchment in Somerset, south-west England (51.08°N and 2.58°W), covering an area of 135 km^2 to its river gauging station at Lovington, is chosen as a case study for the investigations carried out in this work. It is mainly pasture land with some areas of woodland in the higher eastern half of the domain. The choice of catchment is based on the availability of quality data, furthermore the characteristics of the Brue catchment are considered to be representative of rural UK catchments used for rainfall runoff modelling. With six years of continuous data provided by a 49 dense raingauge network (see Figure 1), the Brue catchment provides an ideal study area for the analysis of raingauge network design. The designed scheme can be carefully evaluated with these available dense raingauges. Operationally most catchments are only serviced by two raingauges at best. The cost of maintaining a denser network is too prohibitive for it to be feasible on a larger scale. Clearly the rich data set provided by dense raingauge network is rare, and contains information on rainfall events which would normally be missed by catchments with just one or two raingauges.

Radar and raingauge data sets were maintained by the National Rivers Authority as part of the Hydrological Radar Experiment (HYREX) experiment. The dense raingauge network, radar and a variety of related meteorological data are available through the British Atmospheric Data Centre (BADC). The radar data sets are from the Wardon Hill radar, located at a range around 40 km from the centre of the catchment. The radar completes one cycle through 4 different scan elevations every 5 minutes, and the rainfall intensity is recorded on two Cartesian grids: 5 km and 2 km grid. The raingauges installed on the Brue catchment are typical of those used by the Environmental Agency; a casella tipping bucket gauge mounted vertically on a concrete paving slab. The bucket size was 0.2 mm and the gauge aperture was 400cm^2. The tip time was recorded up to a time resolution of 10 seconds. The first valid day was considered to be the day after which the gauge recorded its first value [*Wood et al.*, 1999]. The dense raingauge network was designed so that all the raingauges would lie entirely in the catchment and that there would be at least one raingauge in the centre of each 2km radar grid square.

Raingauge network design is carried out for 373 rainfall events from September 1993 to April 2000. Except for some events where either radar or gauge data are missing, these events almost cover all significant storms during the period. To display the outcomes, 10 demonstrated typical events are chosen from them. The event ID, durations and averaged rainfall over the catchment are listed in Table 1. The events cover a wide range of scenarios; some of the rainfall are consistent over several hours, (Events 3, 4) and some events are short lived but intense, (Events 2, 9). Averaged rainfall represents the areal averaged rainfall over catchment with accumulated rainfall for an event. It is calculated using radar rainfall measurements. With possible systematic errors, the radar rainfall estimates tend to be larger than gauge rainfall values in Brue catchments. Discussion of radar rainfall adjustment and uncertainty analysis is outside of the scope of this work.

140 The interested readers can refer to our previous papers for more information [*Dai and Han*, 2014;
141 *Dai et al.*, 2014].

142 **3. Methodology**

143 **3.1 PCA application for radar grids network redundancy analysis**

144 PCA is used firstly to examine the redundancy existing in the dense radar grids network and
145 secondly to isolate the grids which provide the most significant contribution to the principal
146 components. The selected radar grids from existing radar grids network are considered to be ideal
147 places of deploying raingauges, which are named optimum grids (OGs). The centre of each
148 optimum grid indicates one potential location for a raingauge. The appropriate number and
149 location of OGs are dependent upon the amount of original variance the network should retain.
150 Principal component analysis (PCA) is a technique used in multivariate analysis where it is
151 suspected that a number of variables are interrelated. It is primarily used for compressing a data
152 set while at the same time minimising the loss of information. This is achieved by generating a
153 new set of variables called principal components (PCs). In this study, redundancy analysis of
154 existing radar grids network based on PCA is used to determine the number of OGs.

155 Given a data set of n variables (i.e., the number of existing radar grids in the study area), with p
156 observations (i.e., the number of hours or days, months etc. observed). The $n \times n$ covariance matrix
157 (C) of the data set is firstly calculated. For PCA to work properly, the original data set is normalized
158 by subtracting the mean from each of the data dimensions. Since the eigenvectors of the covariance
159 matrix (C) are orthogonal, the n eigenvectors can be used as a basis from which the principle
160 components are built, which are shown as follow:

$$Eigenvector = \left(eig_1 \quad eig_2 \quad eig_3 \quad \ldots \quad eig_n\right) \tag{1}$$

Thus the original data can be represented in terms of the *n* eigenvectors via a linear transformation from the original data set *X* to a new data set *Z* where the variance of each of the components is its corresponding eigenvalue:

$$Z_i = X_1 eig_{i,1} + X_2 eig_{i,2} + \ldots + X_n eig_{i,n} \qquad i = 1 \ldots n \tag{2}$$

where X_i and Z_i are the vectors of original and new data sets respectively. These new variables do not contain any redundant information since each principal component is a linear combination of the original variables and all principal components are orthogonal to each other [*Jolliffe*, 1986].

Since the eigenvalue of each component is also its variance, the eigenvector with the highest eigenvalue is the principal component of the data set, this property is used to determine the principal components that carry a set percentage of the variance found in the data. Assessment of the network redundancy is achieved by deciding on a threshold of desired variance explained and then noting how many principal components are required to obtain this threshold. If the threshold is *q*, then we wish to maintain *q*% of variance found in the data, and correspondingly we accept a loss of (1-*q*)% of the original information contained by original radar grids. Thus the number of principal components (say, *k*) required to give us *q*% of explained variance is the number of optimum grids. If *k* is much less than *n*, then we can conclude that the network is heavily redundant and many of the radar grids can be removed without significant loss of information. On contrary, if *k* is quite closely to *n* then there is little redundancy in the network. The choice of the variance threshold is based on two conditions. Basically, the required number *k* obviously grows with the increase of variance threshold. The growth rate will change when the threshold reaches a critical

182 value, which could be regarded as a threshold value. In addition, some subjective factors are also
183 worth considering. For example, the budget-saving-prone or information-remaining-prone
184 designers may have different claims of threshold value under the same condition. The detailed
185 discussion of threshold selection is given in Section 4.3.

186 **3.2 Selection criteria for determining optimum raingauge locations**

187 The principal components are just linear data combinations of all original variables (radar grids
188 rainfalls), so it is necessary to interpret the components in terms of the original variables to select
189 the optimum radar grids (OGs). In effect we can try to choose a subset of the original variables
190 which approximate the information retained in *k* PCs. In this study, cluster analysis (CA) is used
191 to allocate the original variables to a subset of clusters derived from the average linkage method.
192 One variable is retained from each cluster and is chosen as the representative of that cluster. The
193 advantage of CA is that there is no prior knowledge about which elements belong to which clusters.
194 With CA, we can select the OGs and consequently determine the possible locations for raingauges.

195 Formal definition of a cluster, group or class is difficult and is often down to the judgement of the
196 user. *Cormack* [1971] talk of internal cohesion and external isolation in defining clusters. Although
197 there is no standard definition of a cluster it is generally felt that it must have something to do with
198 the recognition of relative distance between members, so certain properties are attributed to
199 clusters such as density, variance, shape and separation which are formed by assessing the
200 similarity and dissimilarity of each pair of members to be clustered. The proximity of each member
201 to the other can be measured in many ways depending on the type of variables under investigation.
202 As we already have concluded the number of clusters (*k* as mentioned above) based on PCA, *k*
203 means clustering is used to partition original *n* variables into *k* clusters in which each variable

belongs to the cluster with the nearest mean, which uses an iterative algorithm to minimize the sum of distances from each object to its cluster centroid, over all clusters. Cluster analysis for selecting OGs works with five main steps: 1) initialize the centroid of the clusters; 2) attribute the closest cluster to each radar grid; 3) set the centroid position of each cluster to the mean of all radar grids belonging to that cluster; 4) repeat steps 2-3 until converges (each centroid stays in a stable location or calculated time reaches the maximum iteration) to form given number of clusters; 5) the OG selected to represent each cluster is chosen in two different ways; the radar grid giving the maximum event-averaged rainfall in a given cluster (CA Max) and the radar grid giving the median of the event-averaged rainfall (CA Med). Cluster analysis can measure distances that are Euclidean (can be measured with a 'ruler') or distances based on similarity. The Euclidean distance is the most straightforward and generally acceptable way of computing distances between objects in a multi-dimensional space. Moreover, the rainfall connection among different radar grids relies on their separated distances. So we adopt the Euclidean distance measure in the proposed scheme. Thus each cluster has one OG and the given number of OGs make up a new compact but high information engaged radar grids network, which brings out a new raingauge network. A flow chart of the proposed method is shown in Figure 2.

3.3 Validation methods

Performance evaluation of raingauge network design is a challenge because no standard assessment criteria exist to indicate what kind of network is the most appropriate one for the given study area. In essence, the designed network should be an effective network, which means the network should contain maximum information at the cost of a minimal number of gauges. This principle can be interpreted by two major rules for this study. First, the selected small number of OGs should maintain the dominating rainfall information of the original radar grids network. And

second, further incensement in the number of OGs should not significantly increase the amount of rainfall information.

The second rule can be achieved through use of an information-component curve, which will be discussed in Section 4.3. For the first rule, two indicators (the Pearson correlation coefficient and Nash-Sutcliffe coefficient) are introduced herein. The Pearson correlation coefficient (r) can estimate the systematic deviation between OGs rainfall (R_m) and the original radar grids rainfall (R_o), which is written as:

$$r_{R_o,R_m} = \frac{E[R_m R_O] - E[R_m]E[R_O]}{\sqrt{\left(E[R_m^2] - E[R_m]^2\right) \times \left(E[R_O^2] - E[R_O]^2\right)}} \quad (3)$$

where E calculates the mean value of the corresponding vector. The Nash-Sutcliffe coefficient [*Nash and Sutcliffe*, 1970] is generally used to assess how well hydrological models predict events. The rainfalls of the designed network and existing network are regarded as modelled and observed values, respectively.

The Nash-Sutcliffe coefficient (NS) is calculated as follows:

$$NS = 1 - \frac{\sum \left(R_o^t - R_m^t\right)^2}{\sum \left(R_o^t - E[R_o]\right)^2} \quad (4)$$

where the superscript t refers to the time-step of the storm and $NS \in [1,-\infty)$. The closer NS is to 1, the more accurate the designed scheme is.

In addition, because radar measurements have non-negligible uncertainty, we speculate that this uncertainty could be propagated to the design processing and contaminate the designed network. Therefore, in the second stage of validation, the less accurate radar data set was compared to the

raingauges data set after implementing the proposed scheme. An existing dense and high-redundancy raingauge network (with 49 raingauges allocated within an area of 135 km^2) was used as a reference network. The proposed scheme was carried out using the raingauge network and radar grids network respectively, and the corresponding gauge-designed as well as radar-designed networks were produced. By comparing the differences of gauge numbers and locations between the gauge-designed and radar-designed networks we can investigate the possible errors of the radar-designed network caused by radar rainfall uncertainty.

The mean number error (E_N) and the mean location error (E_L) are defined to quantitatively describe the designed errors. E_N represents the averaged deviation of sizes between gauge-designed and radar-designed networks, which is derived:

$$E_N = \frac{1}{VN} \sum_{v=1}^{VN} \left| N_O^v - N_m^v \right| \quad (5)$$

where N_O^v and N_m^v refer to the number of gauges in the gauge-designed and radar-designed networks for the given variance threshold v. To avoid the contingency of performance evaluation due to using a certain variance threshold, dozens of threshold values are used to design raingauge networks. VN is the number of involved variance thresholds.

E_L is used to illustrate the disparity in space between two designed networks. The selected radar grids and selected raingauges are first paired and the distances between each pair are accumulated. The combined scheme with the least accumulated distance is adopted and the distance is defined as E_L. If the number of radar grids is smaller than that of raingauges, one radar grid may correspond to two raingauges and vice versa. As the difference of numbers between them is quite small, the error that may be introduced is negligible. The radar-designed network is explicitly optimal if its

gauge locations are exactly the same as these of the gauge-designed network. In such cases, E_L is equal to zero. E_L is calculated for each variance threshold.

4. Model computations and results

4.1 Radar grids network redundancy

With 28 radar grids located within an area of 135 km^2, a redundancy of rainfall should exist in the radar grids network. To show this, the correlation matrixes of 28 radar grids are calculated and drawn in Figure 3 for Events 1 and 2. The Pearson correlation coefficients are computed for every two radar grids. Almost all correlation coefficients in both events are larger than 0.5. For the Event 2, the coefficients are even generally larger than 0.8. The rainfall patterns of the last two radar grids are relatively exceptional compared with other grids' in Event 2. This is explicable as the last two radar grids are located in the southern boundary of the catchment and a bit far away from others. So it is not surprising that the correlations between them and other grids are relatively weak in a certain event. Despite this, two rainfall measurements (one for the southern region and the other for the remaining region) may be enough to represent the rainfall diversity in this event. In summary, the radar grids network could be heavily redundant and the analysis of its redundancy can reveal the optimum number of key locations for deploying raingauges.

PCA applied to the dense radar grids network provides a measurement of the network redundancy for an accepted loss of total information. The analysis is conducted on each storm, and highlights the different event characteristics, which is shown in Figure 4. For most of the events the 1st principal component carries close to 90% of the total variance, with the 2nd component bringing this to over 95% of the total variance. This again indicates a very high level of redundancy in the network, so much so that just 1 component contains 90% of the total information. However two

events, 5 and 6, have a noticeably different component weighting, in the case of Event 5, the first component can account for around 80% of the total information, in order to reach 90% both the first and second components are required. The situation is worse for Event 6, here the first component carries just around 75% of the information, in order to retain 90% of the information, 3 components are required. This shows that there is less redundancy in the network for Events 5 and 6, indicating that the rainfall amount measured in the radar grids is more varied. This suggests that these two events are less uniform than the other 8 events, and so the network will require more radar grids. However, considering 28 radar grids in total, many radar grids are still unessential for these two events. It is worth remarking that the analysis of radar grids network redundancy is served for raingauge network design. The argument of most radar grids in current network is unnecessary does not mean we encourage to apply only part of radar data set in other applications. In reality one of the most important advantages of weather radar is it can take millions of measurements from a single platform and consequently reveal spatial variation of rainfall.

To better show the relationship between principal component numbers and variance explained, thresholds of desired variance explained are set to 75%, 80%, 85%, 90%, 92.5% 95%, 97.5% and 99%. The required number of components are summarised in Table 2. Leaving aside the Events 5 and 6, it can be seen that 2 components are sufficient (although not always necessary) to retain 90% of the information. In addition, to be sure to get at least 95% of the information, then 3 components should be used (although 2 components is sufficient for 6 of the events).

4.2 Location of raingauges

After establishing the level of redundancy in the radar grids network it is necessary to determine which grids to select so that the maximum level of information will be retained and unnecessary

311 repeated measurements would be removed. Since the components do not represent physical radar
312 grids, cluster analysis (CA Max and CA Med) are used.

313 The locations of optimum radar grids (OGs as mentioned above) for each event derived from PCA
314 and CA Max are shown for retaining at least 90% of the total variance, as explained in Figure 5.
315 Each event has a different combination of radar grids and suggests that the optimum locations of
316 a small number of radar grids depend on the rainfall event itself. The selected optimum radar grids
317 tend to locate at the boundary of the catchment, especially for events that claim only one or two
318 grids. For example, in Event 1, two OGs are located at the southern boundary of the catchment,
319 while the OG is chosen from northern boundary in Event 2. The corresponding derived OGs using
320 CA Med method are shown in Figure 6. Inspection of Figures 5 and 6, shows that the different
321 variable selection methods produce different OG locations. In some cases the OG locations differ
322 just slightly (e.g. Event 10), in other cases the locations differ substantially (e.g. Events 8 and 9).

323 It is not practical or cost efficient to choose the most suitable location based on one type of rainfall
324 event. Figures 7 and 8 show the envelope of all events based locations for each method of variable
325 selection. The total number of radar grids needed to satisfy each event is 9 for the CA Max case,
326 which is at least 3 times the requirement of each individual case but still less than half of the
327 original number of 28 radar grids. Figure 7 shows the distribution of the required 9 gauges is
328 predominately at the boundary region of the catchment. The total number of radar grids increases
329 to 14 for the CA Med case (see Figure 8). Figure 8 shows the distribution of the required grids is
330 more evenly spread in contrast to the CA Max selection criterion. These envelopes give an
331 interesting insight into the preferential areas for raingauge location; however, analysis of
332 individual events demonstrates that just 3 radar grids are needed to provide a catchment average

rainfall close to the 28 radar grids catchment average (see Section 4.4), therefore the envelope of all radar grids locations does not lead to an efficient network.

4.3 Comparison of designed networks by radar and gauge data sets

The selection procedure of OGs are carried out for each event, while each event gives different outcomes. In an effort to produce an efficient and reliable network for all events, PCA analysis is repeated on the concatenated set of rainfall data, by concatenating the 373 events that cover a period of six years. The numbers of components to reach the given variance thresholds (75 %, 80%, 85%, 90%, 92.5% 95%, 97.5% and 99%) for concatenated events are shown in Table 3. The 1st principal component carries at least 85% of the total information held in the dense radar grids network. A further 2 components are required to carry 90% of the total information. It can be seen that the 28 radar grids network still has a high level of redundancy. However the number of components required to retain 99% of the total information increased from 4 to 7 for most individual events to at least 12 for the set of concatenated events.

To evaluate the performance of raingauge network design by radar data, a dense raingauge network is also used as a comparison. As mentioned above, the mean number error and the mean location error are two major indicators for the evaluation. The relationships between variances explained and principal components for concatenated events using radar and raingauge data respectively are shown in Figure 9. As the variances explained refer to the information the given number of components can achieve, the relationship is also called information-component curve. It is clearly that the disparity between radar and gauge is quite small. For the variance less than 98%, the differences are no more than 1 component. Moreover, the mean number error calculated using Equation 1 is only 0.59. These facts prove the radar data has similar performance as raingauge data

355 in determining the optimum number of gauges. For both radar and raingauge cases, the required
356 number grows gradually with the increment of the threshold value when the variance threshold is
357 small. From Figure 9, we can observe the inflection point appears at around 95% variance
358 threshold. The required components merely climb from 1 at 84% to 3 at 95% for both radar and
359 gauge cases. Nevertheless, the growth rate is remarkable when the variance is larger than 95%. In
360 other words, if we expect to maintain more than 95% information, only quite limited additional
361 information can be gained while spending numerous additional components. This is obviously
362 cost-inefficient compared to the easily harvest of information when variance is less than 95%. In
363 summary, analysis of information-component curve is a promising method to investigate the
364 efficiency of principal components and guide the determination of variance threshold.

365 The same as individual event, cluster analysis is used to derive the best locations for the raingauges
366 using concatenated events which correspond to different levels of variance explained given by the
367 principal components (85%, 90%, 95% and 99%). Figures 10 displays the locations of the selected
368 radar grids and raingauges by CA Max criterion using radar and gauge data set respectively. The
369 dots represent gauges of the gauge-designed network while boxes illustrates the grids of the radar-
370 designed network. It can be seen from the figure that although selected radar grids (OGs) cannot
371 capture all raingauge points, the distributions of them are quite similar. Take 90% variance for
372 example, two raingauges just locate in the neighbouring grids of selected OGs. The same
373 conclusion can also be drawn from Figure 11, which shows for the CA Med case. One can observe
374 that a raingauge is exactly captured by an OG in 90% variance case. As many gauges and grids
375 are required to retain 99% variance, the corresponding relationship between radar- and gauge-
376 designed networks is not clear in Figures 10 and 11. To quantitatively describe the differences of
377 locations between two networks, the mean location errors of radar-designed network are calculated,

which are shown in Table 4. The centre of OG is used to measure distance towards raingauge, so mean location error still exists even the OG captures the corresponding raingauge. In Table 4, it is delighted to observe the mean location errors are less than 3 km in all cases. For 90% variance using CA Med, the errors are as low as 1.02 km. The averaged values of 1.85 and 1.99 km for CA Max and CA Med respectively also indicate the highly agreement between radar- and gauge-designed networks. So it can be said that the raingauge network design using radar data set can represent gauge data set and in other words, the effect of radar rainfall uncertainty on designing raingauge network using the proposed scheme is inconsiderable.

4.4 Raingauge network design evaluation

As mentioned above, it is important to evaluate whether the radar-designed network can maintain the dominated information of the original radar grids network. Figures 12 and 13 show the scatter plots of each event comparing the 28 radar grids catchment average with the catchment average produced by each of the variable selection method. Overall both methods (CA Max and CA Med) produce the good estimates for average catchment rainfall on an event basis. This is confirmed in Table 5 which gives the corresponding Persons correlation coefficient for each method and event. It is found that all correlation coefficients are larger than 0.90, and even reach 0.99 for Event 6 using CA Max.

The ability to produce dependable catchment average for each individual event was again determined by the Nash-Sutcliffe coefficient and these results are given in Table 6. The values are generally larger than 0.90 for CA Med case and fairly stable for different events. In terms of the CA Max case, we can not only observe events with poor coefficients (e.g., 0.42 of Event 4 and 0.45 of Event 10), but also some high-performance events (e.g., 0.97 of Event 6). A measure of

the range of coefficients given by the interquartile range can provide a clue to the reliability of each method for producing a raingauge network suitable for all tested events (see Figure 14). The greater range occurs with CA Max, which suggests that for this case study this method produces the less reliable network.

Scatter plots of concatenated events comparing the 28 radar grids catchment average are also shown in Figure 15, together with Persons correlation coefficients listed in Table 7. As high as 0.92 and 0.93 of correlation coefficients indicate a fine agreement between radar-designed network and the original radar grids network. In terms of Nash-Sutcliffe coefficient, the one of CA Max is much smaller than that of CA Med (0.69 to 0.84). For this reason, the raingauge network designed by CA Med method can maintain more information than CA Max does in Brue catchment with the same number of OGs.

5. Discussion

A simple, efficient, and quantified-based method is proposed in this study and a series of evaluations proves the good performance of the method. However, there are still some key concerns that readers should be aware of. All methods consistently reduce the dense radar grids network from 28 grids down to 5 or less. The main difficulty is in determining the optimal positioning of the raingauges and the stability of the designed network. The initial study, which analyzed the events individually, produced significant differences in location for each event and method to the point that it was not possible to identify any general pattern. The envelope of all chosen raingauges produced a network capable of detecting the variability of the rainfall field for all events; however, this led to networks of between 9 and 14 radar grids, substantially more than that needed for any one of the events. To tackle this problem, the 373 events covering a period of

6 years were considered as a whole. The proposed scheme was implemented using such concatenated events and the required number of radar grids dropped to three. The locations of the selected OGs given by each selection criterion are more consistent. The analysis of Nash and correlation coefficients between the radar-designed and original network also supports the rationality of this scheme.

In addition, concerns may be raised regarding whether the remotely-sensed rainfall (radar rainfall in this study) uncertainty will influence and contaminate the outcomes of the raingauge network design. It is true that there are numerous, yet to be tackled, problems associated with weather radar, such as ground clutter, anomalous propagation, signal attenuation, beam blockage, and vertical variability of the reflectivity [*Cluckie et al.*, 2000; *Villarini and Krajewski*, 2010]. Many groups have made significant efforts to adjust or describe radar rainfall errors [*Borga et al.*, 2002; *Ciach et al.*, 2007; *Collier*, 1986; *Dai et al.*, 2013; *Dai et al.*, 2015; *Kirstetter et al.*, 2010; *Rico-Ramirez and Cluckie*, 2007; *Villarini et al.*, 2008]. In fact, the overall bias of weather radar will not significantly influence this study. For example, if rainfall values of all radar pixels time a given ratio, the designed network using the proposed scheme will not change. However, the random error of radar will to some extend affects the final results. To evaluate the possible errors of the radar-designed network caused by radar rainfall uncertainty, this study investigated two networks using radar and rain gauge data sets respectively. The small differences of gauge numbers and locations between the gauge-designed and radar-designed networks prove the effective of the proposed scheme. In addition, to reduce the effect of radar rainfall uncertainty on the proposed scheme, we adopted a relatively trustworthy radar data set herein, which is from the Hydrology Radar Experiment (HYREX) conducted by the Natural Environment Research Council (NERC) Special Topic Programme. Typical errors for radar data have been identified and reduced in the initial

445 processing [*Bringi et al.*, 2011; *Moore et al.*, 1999]. Moreover, we used long-term radar rainfall
446 records instead of individual events for the raingauge network design, which could to some extent
447 remove the outliers and produce a more stable outcome. In fact, we propose that a stricter way to
448 solve the uncertainty problem is to integrate the radar rainfall uncertainty model with the proposed
449 scheme. This approach may be necessary in other regions, such as hilly areas where weather radar
450 suffers more problems. The radar rainfall uncertainty model refers to a mathematical approach that
451 elaborately formulates all uncertainties associated with radar rainfall [*Dai et al.*, 2014; *Krajewski*
452 *et al.*, 1991]. An ensemble generation of a large number of probable 'true rainfall' is currently a
453 popular type of radar rainfall uncertainty model [*AghaKouchak et al.*, 2010; *Dai et al.*, 2014;
454 *Germann et al.*, 2009]. For example, we can generate 100 rainfall values that satisfy the error
455 distribution and other restricted conditions of radar rainfall and input them into the proposed
456 scheme to produce 100 possible raingauge networks. Thus, the designed outcome can be expressed
457 in a probabilistic form instead of a determinate network. A decision-making scheme under
458 uncertainty could be introduced to choose the optimum raingauge network.

459 This study makes the assumption that only the center of the radar grid can be used as the potential
460 location of a raingauge. The spatial resolution of radar data used herein is 2 km. Weather radars
461 with higher spatial resolutions such as 1 km or hundreds of meters are becoming increasingly
462 popular all over the world [*Emmanuel et al.*, 2012; *Sandford*, 2015; *Smith et al.*, 2012; *Thorndahl*
463 *et al.*, 2014; *Wright et al.*, 2013]. Considering the natural spatial continuity of rainfall, there should
464 be a distance error tolerance in raingauge network design. The so-called distance error refers to
465 the distance differences between the practical designed raingauge network and the ideal raingauge
466 network that perfectly satisfy the requirements of maximum rainfall information with a minimum
467 number of gauges. For example, in this study where there are no dramatic changes of land surface

terrain, a distance error of hundreds of meters is considered to be acceptable. However, the resolution of satellite rainfall is relatively lower although it increases rapidly. For example, the spatial resolutions of post-processed rainfall products from TRMM and GPM are just 5 km and 4 km, respectively [*Matsui et al.*, 2013; *Simpson et al.*, 1988]. In such situations, scale differences between point raingauges and areal satellite grids may be worth notice. Such point-to-area error has been studied by numerous hydrologists and meteorologists in remotely-sensed rainfall fields and abundant methods of describing or reducing this error have been proposed [*Bringi et al.*, 2011; *Ciach and Krajewski*, 1999; *Habib et al.*, 2004]. Integration of these methods and the proposed scheme offers a promising solution to raingauge network design using satellite rainfall measurements.

6. Conclusions

The results of this study show that PCA combined with simple selection criteria is an effective tool for raingauge network design, and for the given case study. Moreover this new methodology can be used in ungauged catchments as it only requires rainfall data that could be provided by weather radar, satellite or other remotely-sensors. The very nature of PCA is to identify how much information in a data set is useful; this property has been successfully exploited to provide the number of raingauges needed for a chosen level of retained information. The principal components derived from the PCA method do not represent physical raingauges, therefore criteria selection methods are required to identify the best raingauge locations. Cluster analysis methods presented in this study are both simple and quick to use and effective in determining raingauge locations. For individual events the best raingauge locations vary significantly for different events and this means that it is impossible to have an optimum network for individual events unless all the best raingauge locations of all the events are installed. However such a network would be impractical

and expensive to implement. Therefore a compromise must be made based on the concatenated events to derive an overall optimised raingauge network. For the presented case study, two selection criteria (CA Max and CA Med) are used to determine the optimum locations of raingauge network. Both methods can ensure the network designed by radar data set has similar characteristic as that by gauge data set. It is found that CA Max tends to pick the radar grids located at the boundary of catchment, while the selected radar grids from CA Med are distributed more evenly over the catchment. Moreover, CA Med produces a higher performance network from the view of Nash coefficient.

Except cluster analysis, there are also a range of other selection criteria can be used to choose a subset of the original variables which approximate the retained principal components. For example, loading combination criteria (LC), variable deselection (B2) and variable retention (B4) methods [*Al-Kandari and Jolliffe*, 2005; 2007] can also be introduced into the proposed scheme. LC, B2, and B4 selection criteria choose variables according to a given association with the original p variables and the loadings of either the first few components (for variable retention) of or the last (p-q) components, (for variable deselection). Detailed description of these methods are discussed by [*Al-Kandari and Jolliffe*, 2007]. The cluster analysis can be easily replaced by these methods in the proposed scheme. There is no intrinsic disparity among these selection criteria, but they may have different performances in different study areas. One of the major advantages of this study is that the local storm characteristics have already been contained in radar measurements and can be derived through the analysis of long-term radar data. So we believe this study can be easily and effectively extend to other study areas. Since this is the first time that PCA has been used in raingauge network design, we hope more number of study areas with diverse climate and

geographical conditions could be explored by the research community to further verify and improve the proposed scheme.

As said above, hydrologists and meteorologists are facing a severely shut of the available raingauges [*Overeem et al.*, 2013]. So the proposed scheme can not only be used for raingauge network design in an ungauged catchment, but also help reducing the current gauges in a reasonable form. Since the emergence of remotely-sensed rainfall measurements, raingauge continually playing a giving role that assists remotely-sensors to offer a more trustworthy rainfall measurements. This study is a preliminary attempt that using remotely-sensors data set to solve the traditional raingauge problems. Based on remotely-sensed rainfall information, other problems such as wind effect on raingauge may also be studied in a new insight and the original complicated problems may be solved in a simple scheme just as what have been discussed in this study.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 41501429), the PAPD program (Grant No. 164320H116) and the University Natural Science Project of Jiangsu Province (Grant No. 16KJA170001). The authors acknowledge the British Atmospheric Data Centre for providing the data.

References

AghaKouchak, A., A. Bárdossy, and E. Habib (2010), Copula‐based uncertainty modelling: application to multisensor precipitation estimates, *Hydrological processes*, *24*(15), 2111-2124.

Al-Kandari, N. M., and I. T. Jolliffe (2005), Variable selection and interpretation in correlation principal components, *Environmetrics*, *16*(6), 659-672.

Al-Kandari, N. M., and I. T. Jolliffe (2007), Variable selection and interpretation of covariance principal components, *Communications in Statistics - Simulation and Computation*, *30*(2), 339-354.

Al-Zahrani, M., and T. Husain (1998), An algorithm for designing a precipitation network in the south-western region of Saudi Arabia, *Journal of Hydrology*, *205*(3), 205–216.

Barancourt, C., J. Creutin, and J. Rivoirard (1992), A method for delineating and estimating rainfall fields, *Water Resources Research*, *28*(4), 1133-1144.

Bastin, G., B. Lorent, C. Duqué, and M. Gevers (1984), Optimal Estimation Of The Average Areal Rainfall And Optimal Selection Of Rain Gauge Locations, *Water Resources Research*, *20*(4), 463–470.

Bogárdi, I., A. Bárdossy, and L. Duckstein (1985), Multicriterion Network Design Using Geostatistics, *Water Resources Research*, *21*(2), 199–208.

Borga, M., F. Tonelli, R. J. Moore, and H. Andrieu (2002), Long‐term assessment of bias adjustment in radar rainfall estimation, *Water Resources Research*, *38*(11), 8-1-8-10.

Bradley, A. A., C. Peters-Lidard, B. R. Nelson, J. A. Smith, and C. B. Young (2002), Rainage Network Design Using Nexrad Precipitation Estimates, *Journal of the American Water Resources Association*, *38*, 1393-1407.

Bras, R. L., and I. Rodríguez-Iturbe (1985), Random Functions and Hydrology, *Random Functions & Hydrology*.

Bringi, V., M. Rico-Ramirez, and M. Thurai (2011), Rainfall estimation with an operational polarimetric C-band radar in the United Kingdom: comparison with a gauge network and error analysis, *Journal of Hydrometeorology*, *12*(5), 935-954.

Ciach, G. J., and W. F. Krajewski (1999), On the estimation of radar rainfall error variance, *Advances in Water Resources*, *22*(6), 585-595.

Ciach, G. J., W. F. Krajewski, and G. Villarini (2007), Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data, *Journal of Hydrometeorology*, *8*(6), 1325-1347.

Cluckie, I., R. Griffith, A. Lane, and K. Tilford (2000), Radar hydrometeorology using a vertically pointing radar, *Hydrology and Earth System Sciences Discussions*, *4*(4), 565-580.

Collier, C. (1986), Accuracy of rainfall estimates by radar, Part I: Calibration by telemetering raingauges, *Journal of Hydrology*, *83*(3), 207-223.

Cormack, R. M. (1971), A review of classification, *Journal of the Royal Statistical Society. Series A (General)*, 321-367.

Dai, Q., and D. Han (2014), Exploration of discrepancy between radar and gauge rainfall estimates driven by wind fields, *Water Resources Research*, *50*(11), 8571-8588.

Dai, Q., D. Han, M. A. Rico-Ramirez, and T. Islam (2013), The impact of raindrop drift in a three-dimensional wind field on a radar–gauge rainfall comparison, *International journal of remote sensing*, *34*(21), 7739-7760.

Dai, Q., D. Han, M. Rico-Ramirez, and P. K. Srivastava (2014), Multivariate distributed ensemble generator: A new scheme for ensemble radar precipitation estimation over temperate maritime climate, *Journal of Hydrology*, *511*(7), 17–27.

576 Dai, Q., M. A. Rico‐Ramirez, D. Han, T. Islam, and S. Liguori (2015), Probabilistic radar
577 rainfall nowcasts using empirical and theoretical uncertainty models, *Hydrological*
578 *Processes*, *29*(1), 66-79.

579 Emmanuel, I., H. Andrieu, and P. Tabary (2012), Evaluation of the new French operational
580 weather radar product for the field of urban hydrology, *Atmospheric Research*, *103*, 20-32.

581 Germann, U., M. Berenguer Ferrer, D. Sempere Torres, and M. Zappa (2009), REAL-Ensemble
582 radar precipitation estimation for hydrology in a mountainous region, *Quarterly Journal of*
583 *the Royal Meteorological Society*, *135*, 445-456.

584 Habib, E., G. J. Ciach, and W. F. Krajewski (2004), A method for filtering out raingauge
585 representativeness errors from the verification distributions of radar and raingauge rainfall,
586 *Advances in Water Resources*, *27*(10), 967-980.

587 Bradley, A. A., C. Peters-Lidard, B. R. Nelson, J. A. Smith, and C. B. Young (2007), Raingage
588 Network Design Using NEXRAD Precipitation Estimates, *Jawra Journal of the American*
589 *Water Resources Association*, *38*(5), 1393–1407.

590 Jolliffe, I. T. (1986), *Principal component analysis*, Springer Berlin.

591 Kirstetter, P.-E., G. Delrieu, B. Boudevillain, and C. Obled (2010), Toward an error model for
592 radar quantitative precipitation estimation in the Cévennes–Vivarais region, France, *Journal*
593 *of Hydrology*, *394*(1), 28-41.

594 Krajewski, W. F., V. Lakshmi, K. P. Georgakakos, and S. C. Jain (1991), A Monte Carlo study
595 of rainfall sampling effect on a distributed catchment model, *Water resources research*,
596 *27*(1), 119-128.

597 Krstanovic, P. F., and V. P. Singh (1992), Evaluation of rainfall networks using entropy: I.
598 Theoretical development, *Water Resources Management*, *6*(4), 279-293.

599 Lorenz, C., and H. Kunstmann (2012), The hydrological cycle in three state-of-the-art
600 reanalyses: intercomparison and performance analysis, *Journal of Hydrometeorology*,
601 *13*(5), 1397-1420.

602 Matsui, T., T. Iguchi, X. Li, M. Han, W.-K. Tao, W. Petersen, T. L'Ecuyer, R. Meneghini, W.
603 Olson, and C. D. Kummerow (2013), GPM satellite simulator over ground validation sites,
604 *Bulletin of the American Meteorological Society*, *94*(11), 1653-1660.

605 Mishra, A. K., and P. Coulibaly (2009), Developments in hydrometric network design: A review,
606 *Reviews of Geophysics*, *47*(2).

607 Moore, R. J., D. A. Jones, D. R. Cox, and V. S. Isham (1999), Design of the HYREX raingauge
608 network, *Hydrology & Earth System Sciences*, *4*(4), 521-530.

609 Morrissey, M. L., J. A. Maliekal, J. S. Greene, and J. Wang (1995), The Uncertainty Of Simple
610 Spatial Averages Using Rain Gauge Networks, *Water Resources Research*, *31*(8), 2011–
611 2017.

612 Moss, M. E., and G. D. Tasker (1991), An intercomparison of hydrological network-design
613 technologies, *Hydrological Sciences Journal/journal Des Sciences Hydrologiques*, *36*.

614 Nash, J., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I—A
615 discussion of principles, *Journal of hydrology*, *10*(3), 282-290.

Overeem, A., H. Leijnse, and R. Uijlenhoet (2013), Country-wide rainfall maps from cellular communication networks, *Proceedings of the National Academy of Sciences*, *110*(8), 2741-2745.

Pardo-Igúzquiza, E. (1998), Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing, *Journal of Hydrology*, *210*(1), 206-220.

Rico-Ramirez, M., and I. Cluckie (2007), Bright‐band detection from radar vertical reflectivity profiles, *International Journal of Remote Sensing*, *28*(18), 4013-4025.

Rodríguez-Iturbe, I., and J. M. Mejía (1974), The Design of Rainfall Networks in Time and Space, *Water Resources Research*, *10*(4), 713–728.

Sandford, C. (2015), Correcting for wind drift in high resolution radar rainfall products: a feasibility study, *Journal of Hydrology*.

Shih, S. F. (1982), Rainfall variation analysis and optimization of gaging systems, *Water Resources Research*, *18*(4), 1269–1277.

Simpson, J., R. F. Adler, and G. R. North (1988), A proposed tropical rainfall measuring mission (TRMM) satellite, *Bulletin of the American meteorological Society*, *69*(3), 278-295.

Smith, J. A., M. L. Baeck, G. Villarini, C. Welty, A. J. Miller, and W. F. Krajewski (2012), Analyses of a long‐term, high‐resolution radar rainfall data set for the Baltimore metropolitan region, *Water Resources Research*, *48*(4).

Stedinger, J. R., and G. D. Tasker (1985), Regional Hydrologic Analysis 1. Ordinary, Weighted, and Generalized Least Squares Compared, *Water Resources Research*, *21*(9), 1421–1432.

Tasker, G. D., and M. E. Moss (1979), Analysis of Arizona Flood Data Network for regional information, *Water Resources Research*, *15*(6), 1791-1796.

Thorndahl, S., J. E. Nielsen, and M. R. Rasmussen (2014), Bias adjustment and advection interpolation of long-term high resolution radar rainfall series, *Journal of Hydrology*, *508*, 214-226.

Tsintikidis, D., K. P. Georgakakos, J. A. Sperfslage, D. E. Smith, and T. M. Carpenter (2011), Precipitation Uncertainty and Raingauge Network Design within Folsom Lake Watershed, *American Society of Civil Engineers*, *7*(2), 175-184.

Villarini, G., and W. F. Krajewski (2010), Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall, *Surveys in Geophysics*, *31*(1), 107-129.

Villarini, G., P. V. Mandapaka, W. F. Krajewski, and R. J. Moore (2008), Rainfall and sampling uncertainties: A rain gauge perspective, *Journal of Geophysical Research: Atmospheres (1984–2012)*, *113*(D11).

Volkmann, T. H., S. W. Lyon, H. V. Gupta, and P. A. Troch (2010), Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain, *Water resources research*, *46*(11).

Walsh, D. (2012), The tricky business of counting rain, *NY Times*.

655 Wood, S., D. Jones, and R. Moore (1999), Accuracy of rainfall measurement for scales of
656 hydrological interest, *Hydrology and Earth System Sciences*, *4*(4), 531-543.

657 Wright, D. B., J. A. Smith, G. Villarini, and M. L. Baeck (2013), Estimating the frequency of
658 extreme rainfall using weather radar and stochastic storm transposition, *Journal of*
659 *hydrology*, *488*, 150-165.

660 Yang, Y., and D. Burn (1994), An entropy approach to data collection network design, *Journal*
661 *of Hydrology*, *157*(94), 307–324.

662 Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitoh (2012),
663 APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based
664 on a dense network of rain gauges, *Bulletin of the American Meteorological Society*, *93*(9),
665 1401-1415.

666

667

Figure 1. The Brue catchment terrain map. The cyan dots represent the raingauges and the grid represents the real radar grids. The labelled number is the index of the corresponding grid.

Figure 2. Flow chart of the proposed method.

Figure 3. Correlation matrixes for Events 1 (left) and 2 (right) with each pixel representing a correlation coefficient between the grid indexed from left to right and the grid indexed from up to down.

Figure 4. Variances explained by the principal components for the 10 typical events

Figure 5. Locations of the optimum radar grids (OGs) derived from the CA Max method for the 10 typical events (total variance explained ≥90).

Figure 6. Locations of the optimum radar grids (OGs) derived from the CA Med method for the 10 typical events (total variance explained ≥90).

Figure 7. Locations of the optimum radar grids satisfying all typical events. Derived from PCA and CA Max method (total variance explained ≥90).

Figure 8. Locations of the optimum radar grids satisfying all typical events. Derived from PCA and CA Med method (total variance explained ≥90).

Figure 9. Relationships between the variance explained and principal components for concatenated events derived by radar and raingauge datasets.

Figure 10. Comparison of radar grids and raingauges from radar- and gauge-designed networks with variance thresholds of 85%, 90%, 95% and 99% using the CA Max method. The subtitle 'Var' refers to the corresponding variance threshold.

Figure 11. Comparison of radar grids and raingauges from radar- and gauge-designed networks with variance thresholds of 85%, 90%, 95% and 99% using the CA Med method.

Figure 12. Rainfall correlations between the original 28 radar grids and the radar-deigned network using the CA Max method.

Figure 13. Rainfall correlations between the original 28 radar grids and the radar-deigned network using the CA Med method.

694 **Figure 14.** The range of Nash Coefficients for the two selection criteria applied on 10 typical
695 events.

696 **Figure 15.** Rainfall correlations between the original 28 radar grids and the reduced radar grids
697 of the radar-deigned network using the CA Max (left) and CA Med (right) methods for
698 concatenated events.

699

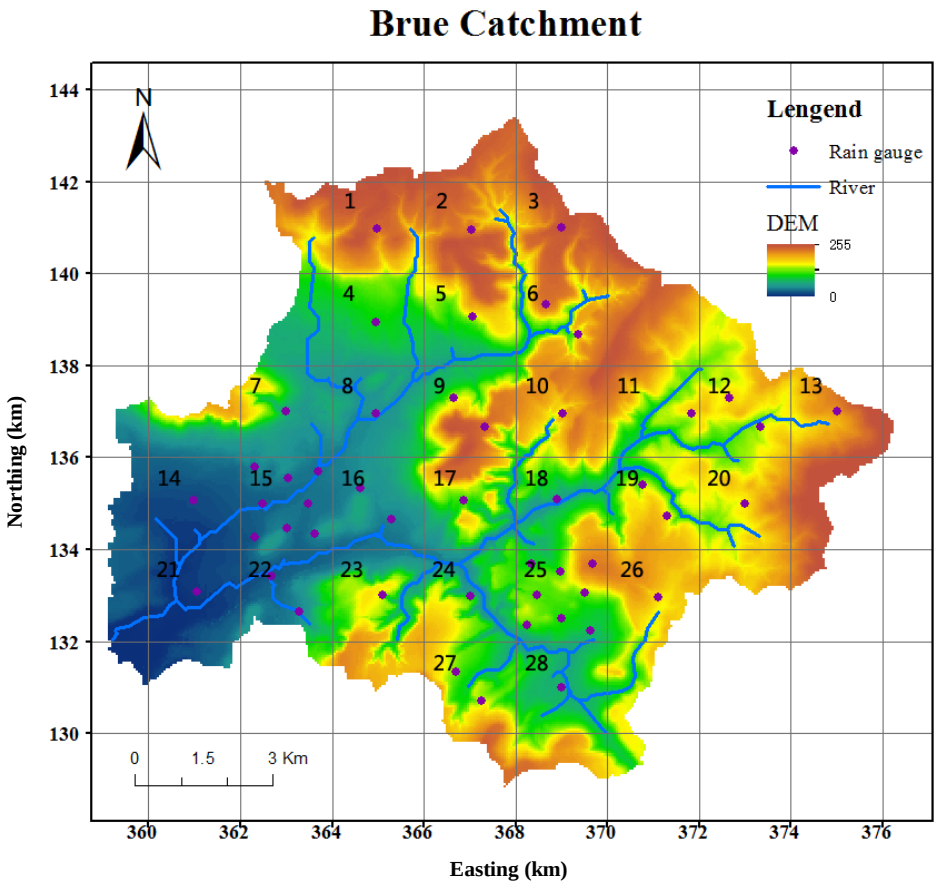
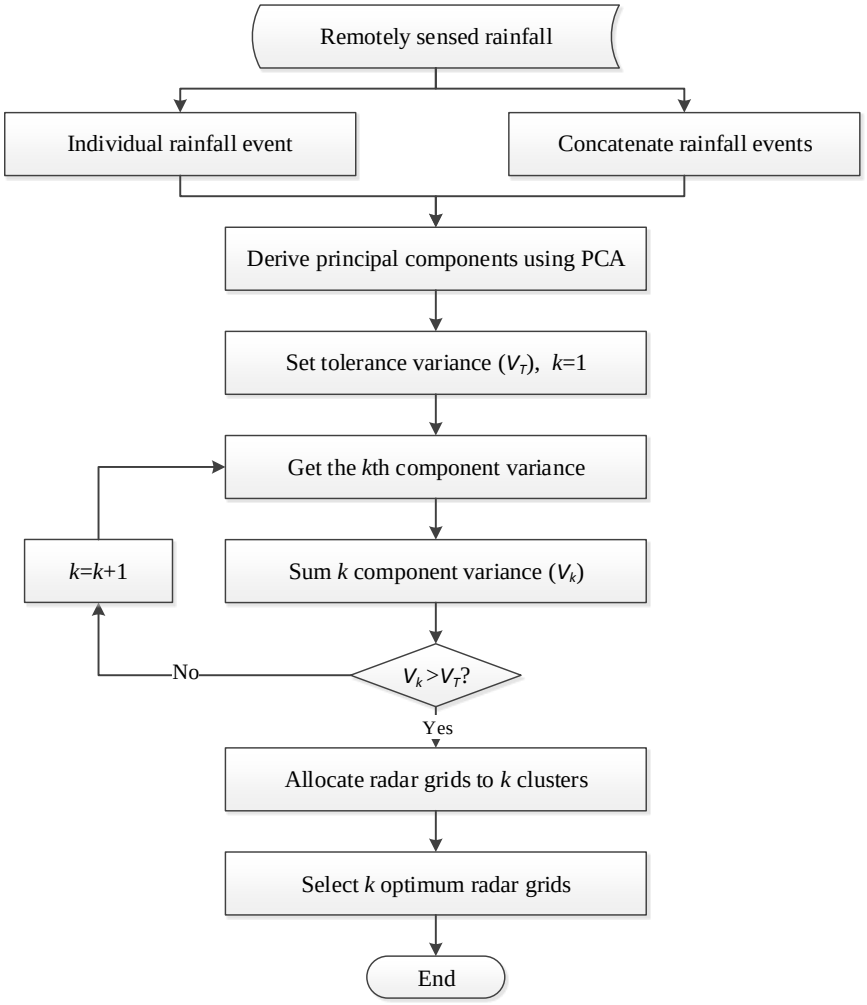


Figure 1. The Brue catchment terrain map. The cyan dots represent the raingauges and the grid represents the real radar grids. The labelled number is the index of the corresponding grid.



704

705 **Figure 2.** Flow chart of the proposed method.

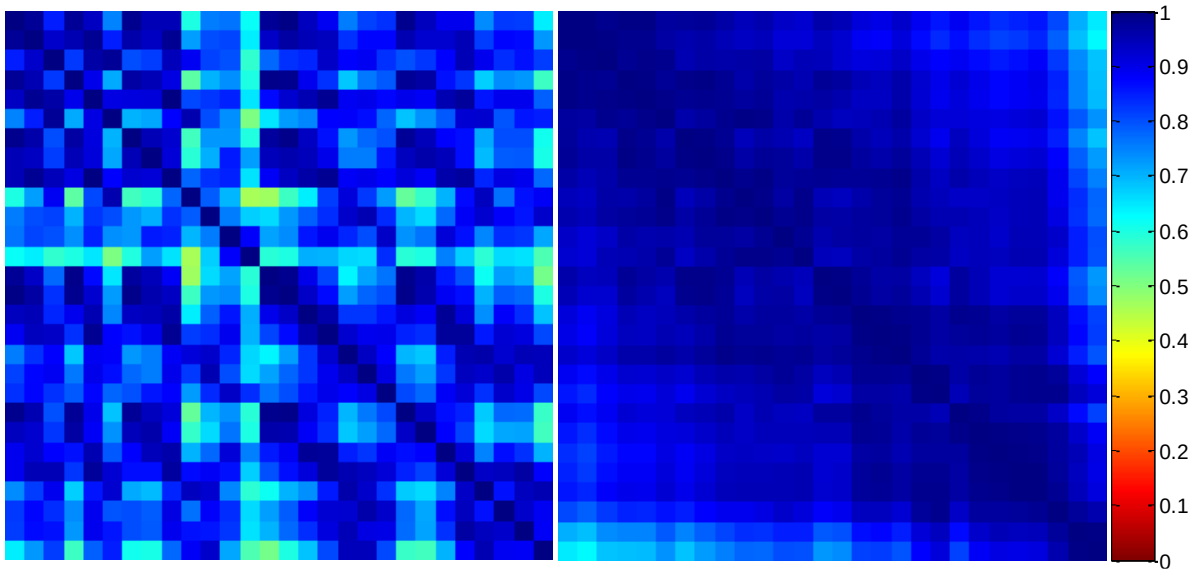
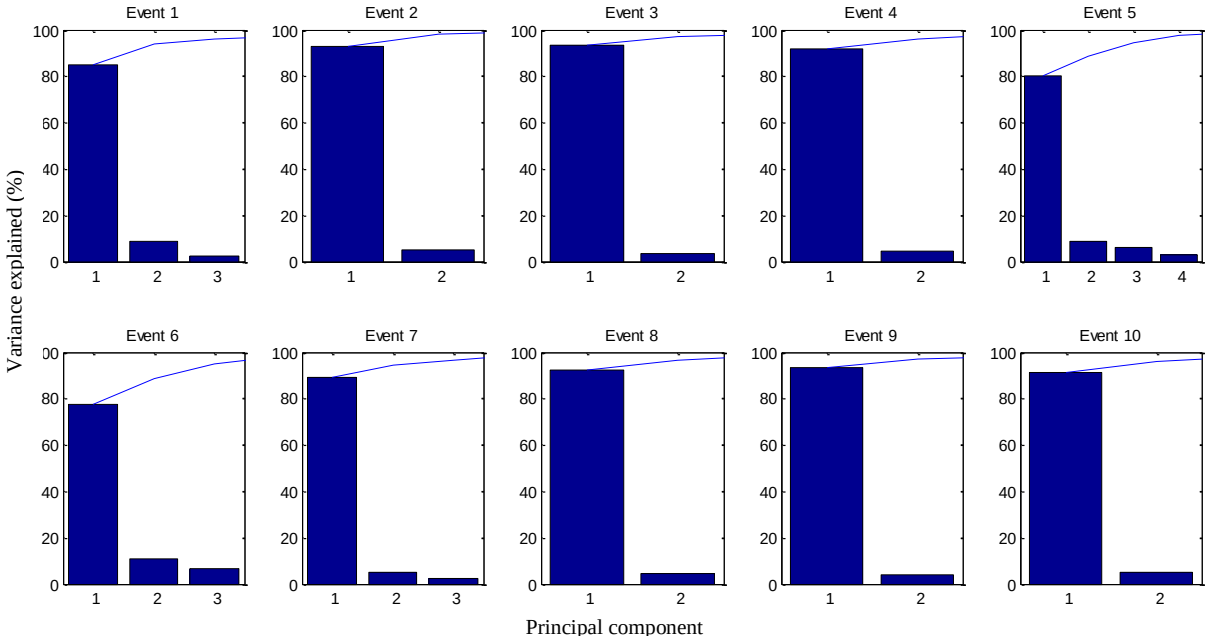
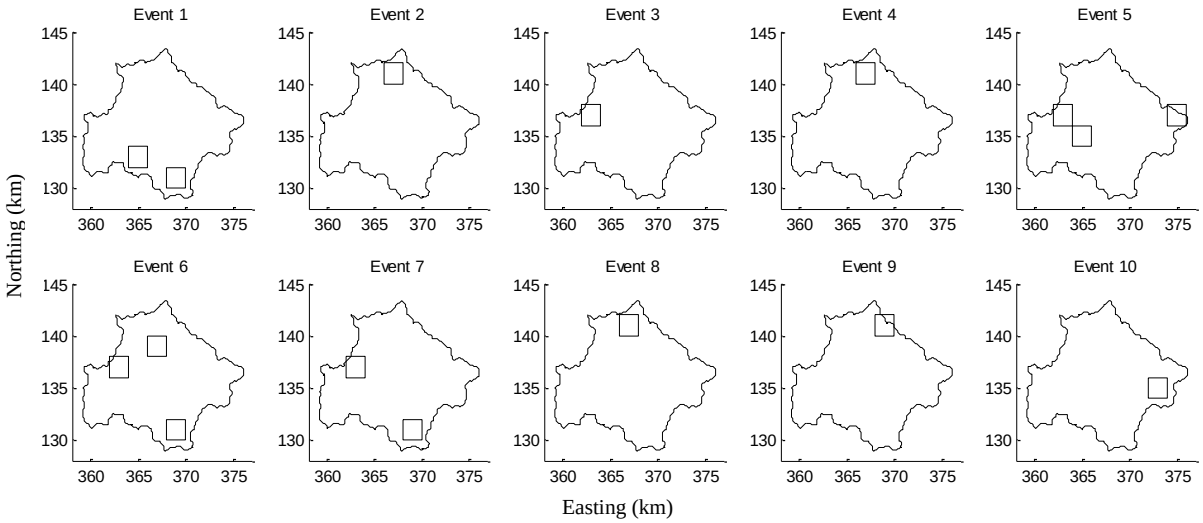


Figure 3. Correlation matrixes for Events 1 (left) and 2 (right) with each pixel representing a correlation coefficient between the grid indexed from left to right and the grid indexed from up to down.



710

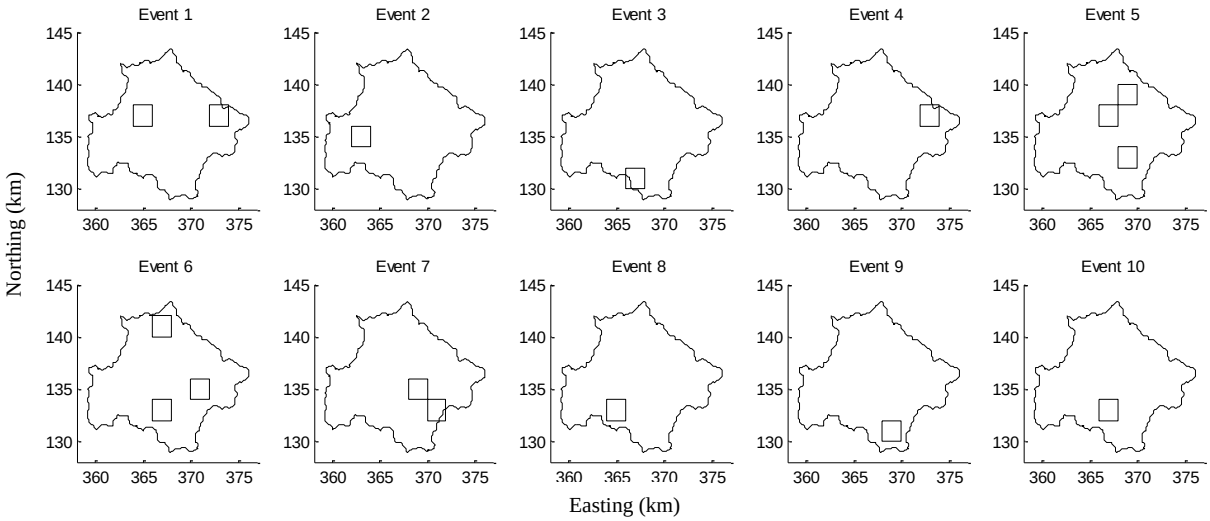
711 **Figure 4.** Variances explained by the principal components for the 10 typical events



712

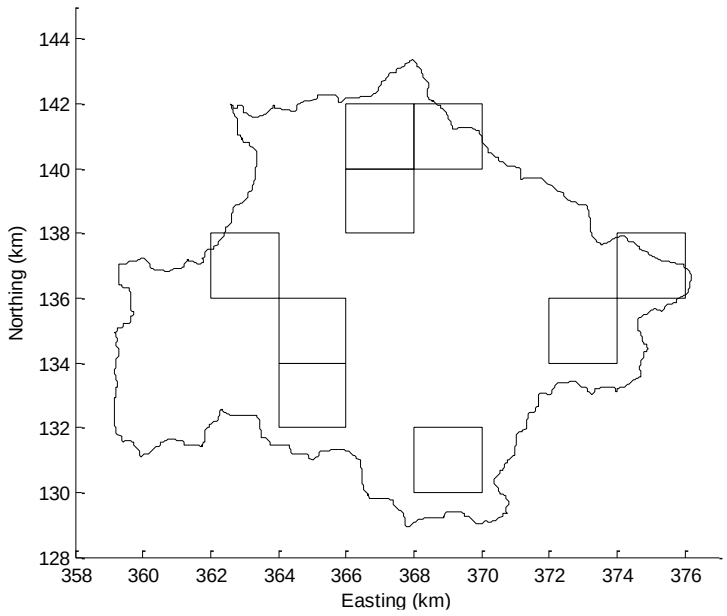
713 **Figure 5.** Locations of the optimum radar grids (OGs) derived from the CA Max method for the
714 10 typical events (total variance explained ≥90).

715



716

717 **Figure 6.** Locations of the optimum radar grids (OGs) derived from the CA Med method for the
718 10 typical events (total variance explained ≥90).



719

720 **Figure 7.** Locations of the optimum radar grids satisfying all typical events. Derived from PCA
721 and CA Max method (total variance explained ≥90).

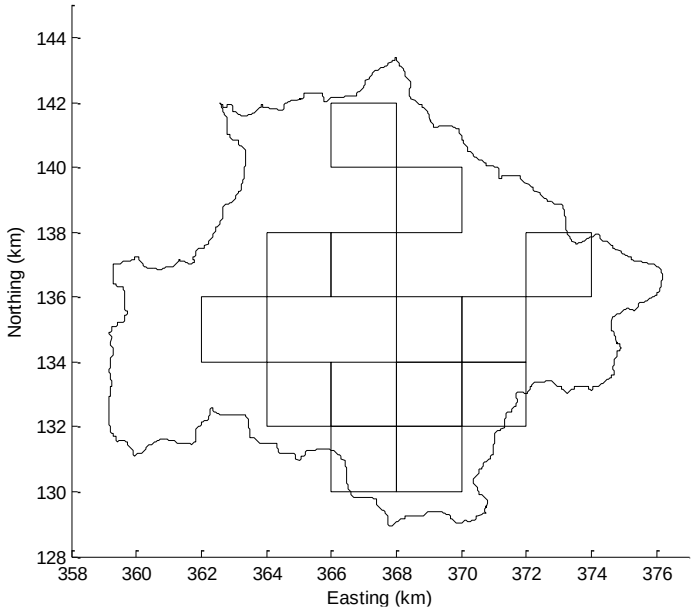


Figure 8. Locations of the optimum radar grids satisfying all typical events. Derived from PCA and CA Med method (total variance explained ≥90).

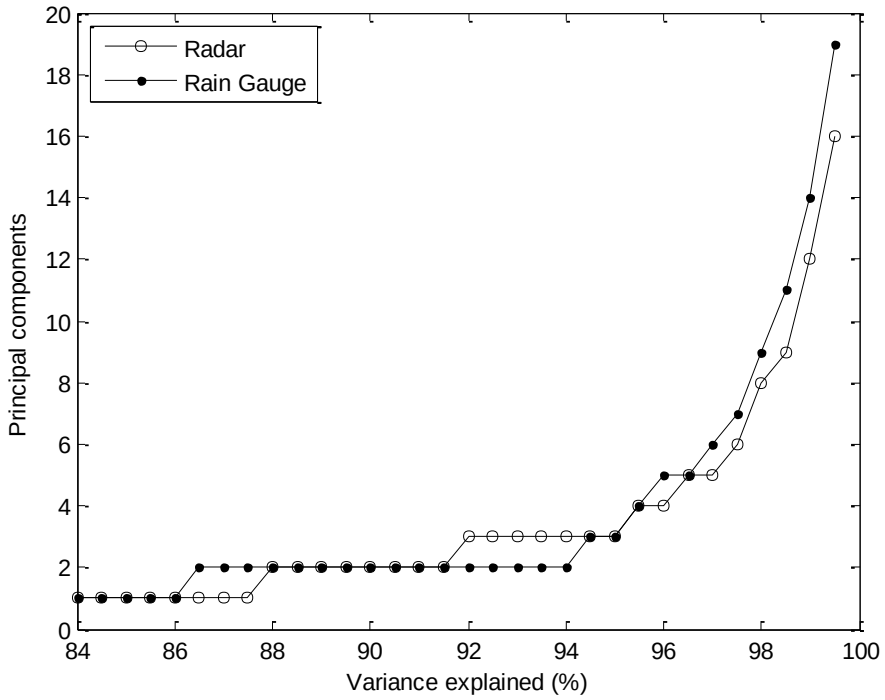
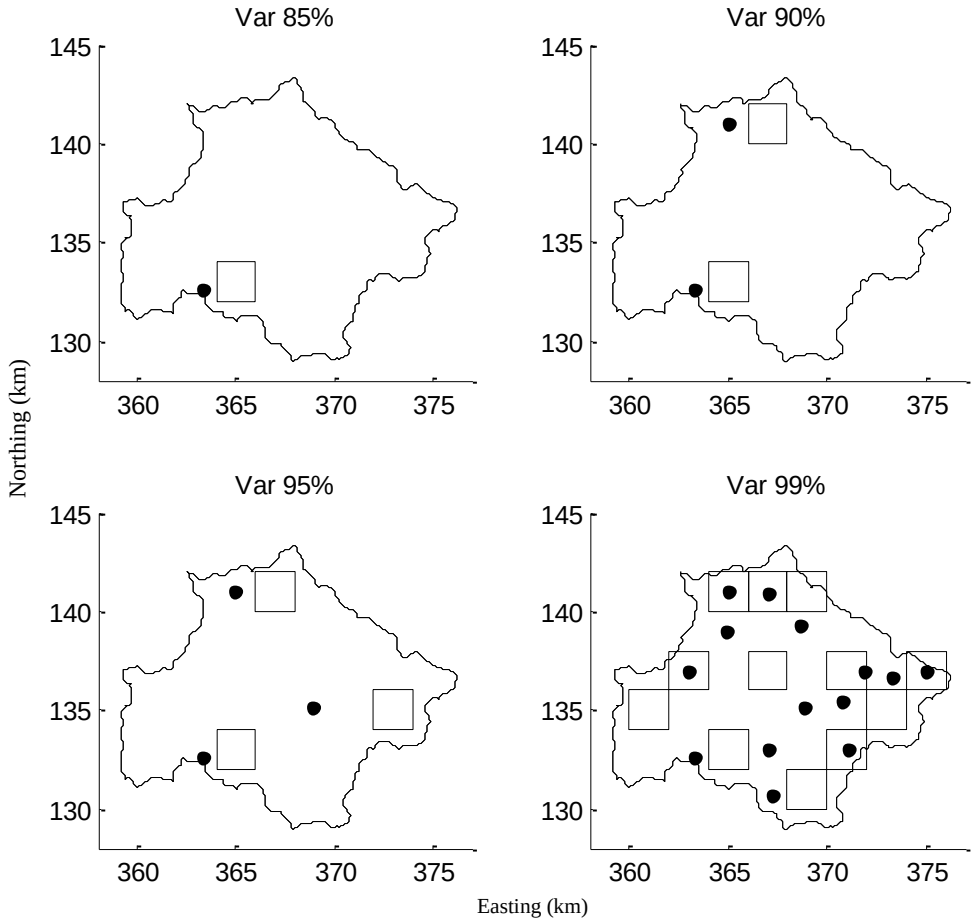
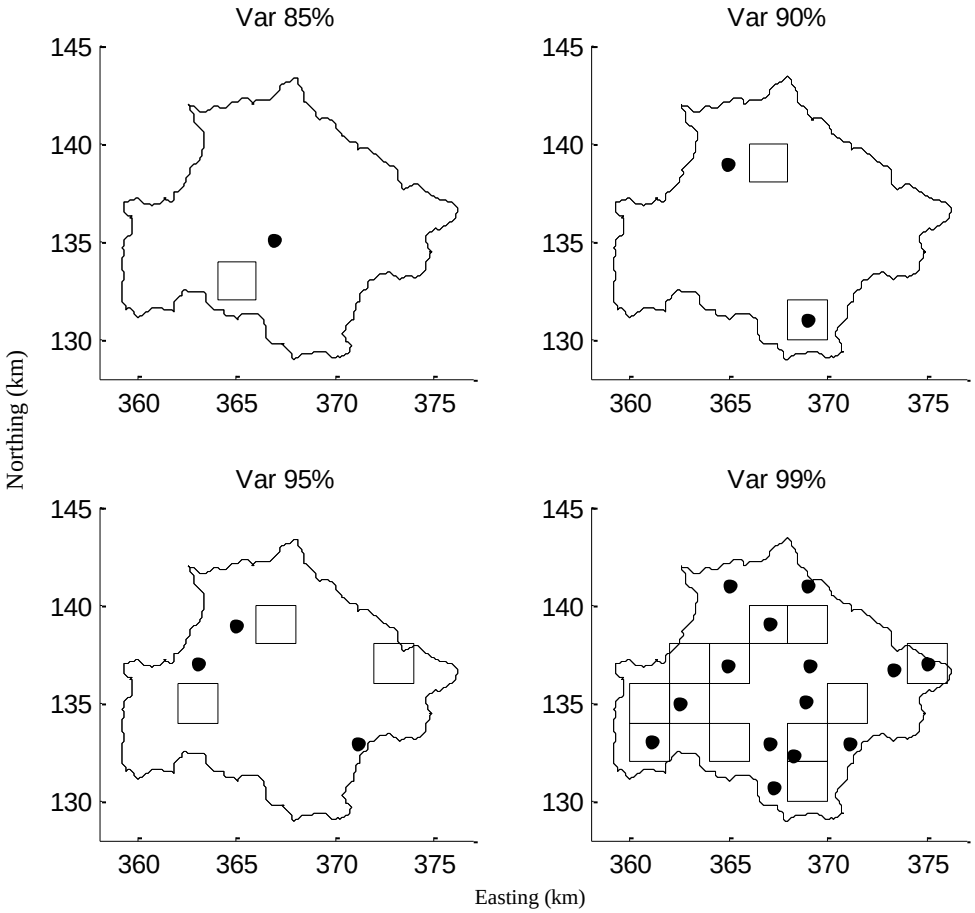


Figure 9. Relationships between the variance explained and principal components for concatenated events derived by radar and raingauge datasets.



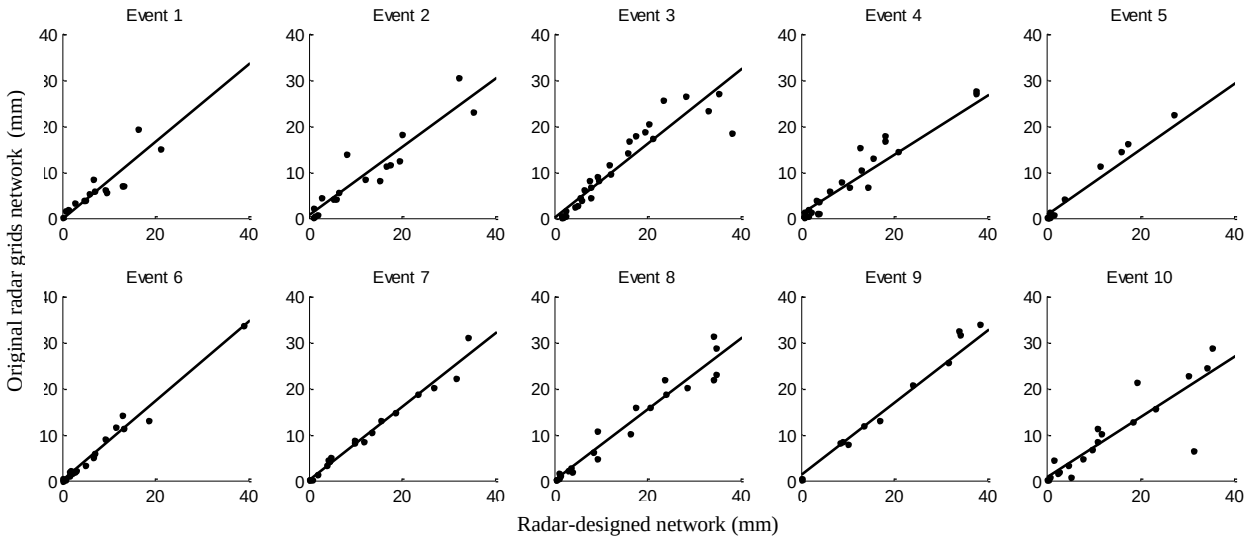
728

729 **Figure 10.** Comparison of radar grids and raingauges from radar- and gauge-designed networks
730 with variance thresholds of 85%, 90%, 95% and 99% using the CA Max method. The subtitle 'Var'
731 refers to the corresponding variance threshold.



732

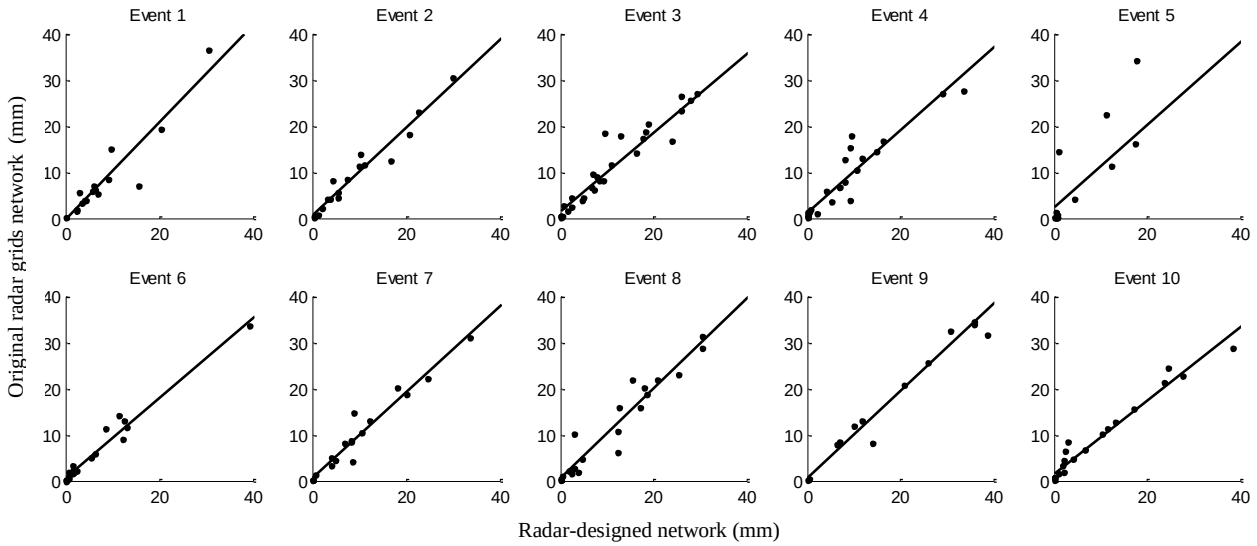
733 **Figure 11.** Comparison of radar grids and raingauges from radar- and gauge-designed networks
734 with variance thresholds of 85%, 90%, 95% and 99% using the CA Med method.



735

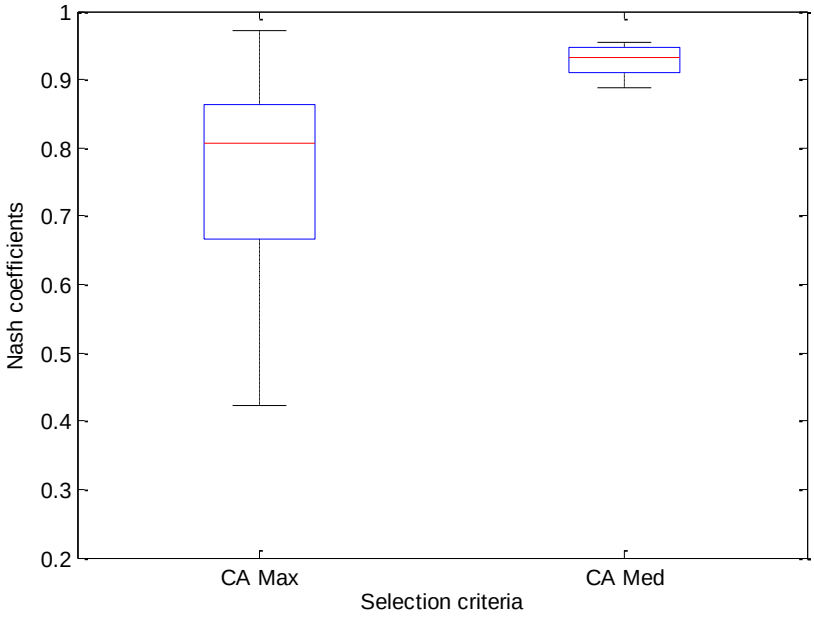
736 **Figure 12.** Rainfall correlations between the original 28 radar grids and the radar-deigned network
737 using the CA Max method.

738



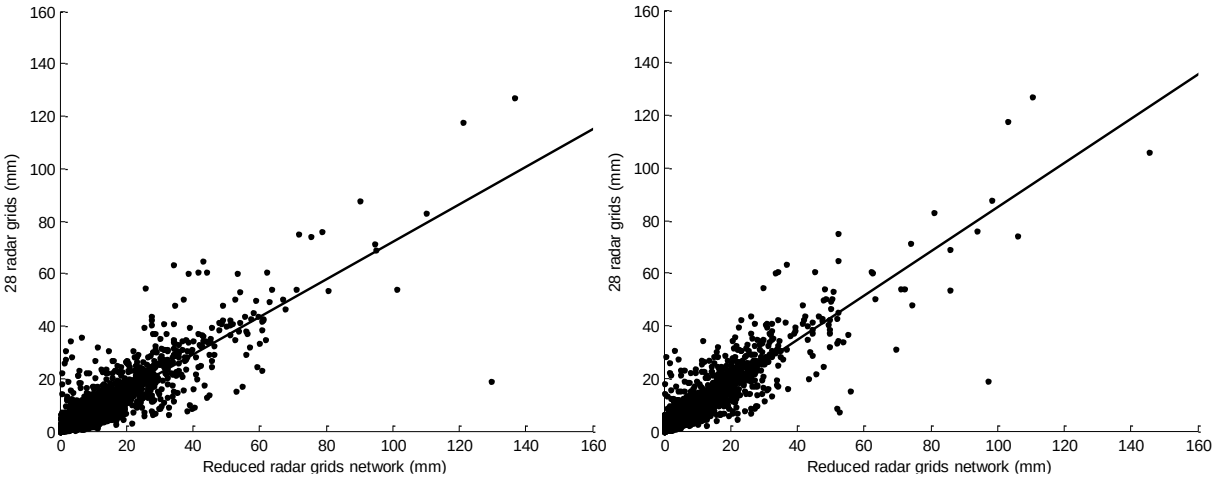
739 **Figure 13.** Rainfall correlations between the original 28 radar grids and the radar-deigned network
740 using the CA Med method.

741



742

743 **Figure 14.** The range of Nash Coefficients for the two selection criteria applied on 10 typical
744 events.



745

746 **Figure 15.** Rainfall correlations between the original 28 radar grids and the reduced radar grids of
747 the radar-deigned network using the CA Max (left) and CA Med (right) methods for concatenated
748 events.

749 **Table 1.** 10 typical rainfall events measured over the Brue catchment. The averaged rainfall
750 represents the accumulated radar rainfall for an event and areal averaged rainfall in the catchment.

Event ID	Start date	End date	Averaged Rainfall (mm)
1	05/10/1993 05:00	05/10/1993 22:00	126.39
2	30/12/1993 02:00	30/12/1993 18:00	153.45
3	08/11/1994 14:00	09/11/1994 22:00	334.51
4	27/01/1995 08:00	28/01/1995 08:00	230.02
5	16/07/1995 11:00	17/07/1995 03:00	150.68
6	26/11/1995 14:00	27/11/1995 10:00	115.41
7	11/02/1996 20:00	12/02/1996 14:00	167.29
8	24/10/1998 04:00	25/10/1998 00:00	230.53
9	18/12/1999 12:00	19/12/1999 01:00	260.52
10	18/04/2000 05:00	19/04/2000 04:00	180.65

751

752 **Table 2.** The number of components to reach % variance threshold for individual events.

Variance	1	2	3	4	5	6	7	8	9	10
75.0	1	1	1	1	1	1	1	1	1	1
80.0	1	1	1	1	1	2	1	1	1	1
85.0	1	1	1	1	2	2	1	1	1	1
90.0	2	1	1	1	3	3	2	1	1	1
92.5	2	1	1	2	3	3	2	2	1	2
95.0	3	2	2	2	4	3	3	2	2	2
97.5	4	2	3	3	5	4	4	3	3	3
99.0	7	4	4	5	5	5	6	4	4	5

753

754 **Table 3.** The number of components to reach % variance threshold for concatenated events

Variance	Component
75.0	1
80.0	1
85.0	1
90.0	2
92.5	3
95.0	3
97.5	6
99.0	12

755

Table 4. Mean location errors (km) for different variance thresholds derived by the two selection methods.

Method	85%	90%	95%	99%	Averaged
CA Max	1.74	1.87	2.62	1.17	1.85
CA Med	2.80	1.02	2.83	1.29	1.99

759 **Table 5.** Correlation coefficients between the 28 radar grids network and the radar-designed
760 network for individual events.

Method	1	2	3	4	5	6	7	8	9	10
CA Max	0.97	0.93	0.95	0.91	0.98	0.99	0.98	0.97	0.98	0.90
CA Med	0.94	0.98	0.97	0.96	0.98	0.98	0.97	0.96	0.98	0.98

761

762 **Table 6.** Nash-Sutcliffe coefficients between the 28 radar grids network and radar-designed
763 network for individual event.

Method	1	2	3	4	5	6	7	8	9	10
CA Max	0.86	0.67	0.80	0.42	0.86	0.97	0.73	0.75	0.82	0.45
CA Med	0.89	0.95	0.91	0.92	0.94	0.95	0.95	0.92	0.95	0.90

764

765 **Table 7.** Correlation coefficients and Nash-Sutcliffe coefficients between the 28 radar grids
766 network and radar-designed network for concatenated events

Method	Correlation	Nash-Sutcliffe
CC Max	0.92	0.69
CC Med	0.93	0.84

767